Appendix G

Trend Analysis of Regional Surface Waters

TREND ANALYSIS METHODS

Trend analyses were performed to assess the historic conditions of near- and far-field pollutant concentrations in the waters upstream and downstream of the Sacramento Regional Wastewater Treatment Plant (SRWTP). The analysis was performed to determine whether any upward or downward trends in concentration over time exist in these waters. The following discussion provides information about the trend analysis methods and an interpretation of the results. The results of the trend analyses are presented in the individual pollutant evaluations for Category 1 constituents included in the antidegradation analysis.

The following describes the way in which the trend analyses were performed. After an initial review of the data (described in Appendix F: Data Sources and Data Quality Screening), distribution testing was performed to determine the normality of each dataset. Necessary dataset adjustments were made and statistical regression analyses were performed on normal or log-normal datasets, as described below.

DATA DISTRIBUTION TESTING

Frequency distribution (probability) plots were created for each constituent at each location evaluated. An example of such a probability plot is shown in **Figure 1**. Probability plots illustrate the distribution of concentrations and allow assessment of how well the data fit a "normal" distribution. The plot provides information on the symmetry of the concentration data relative to a "normal" (i.e., bell-shaped curve) or "log-normal" (single-tailed curve) probability distribution. On such a plot, an evenly-distributed set of data points along a regression line with an R^2 value of 1.0 would represent an ideal normal distribution.

DATASET ADJUSTMENTS

Log-normally distributed data were log-transformed prior to further analysis. All non-detected results were set equal to their reporting limits. However, datasets with large percentages of non-detected results were typically not normally or log-normally distributed, and therefore were not used to perform statistical regression analyses. Chronologically speaking, when the early portion of a constituent's dataset contained very high reporting limits (higher than most subsequently detected concentrations), those data were removed from the dataset, as they would result in the appearance of a false trend. When a dataset contained multiple reporting limits within a close range of values, the reporting limits were *equalized* (i.e., set equal to the median value of the various reporting limits) to prevent the similar appearance of a false trend. Sacramento River flow data, where available, were matched with concentration data for normally and log-normally distributed datasets at locations along the Sacramento River. An equation was developed to model seasonal changes and paired with every normally or log-normally distributed dataset. All dates were converted to decimal dates for statistical analysis.

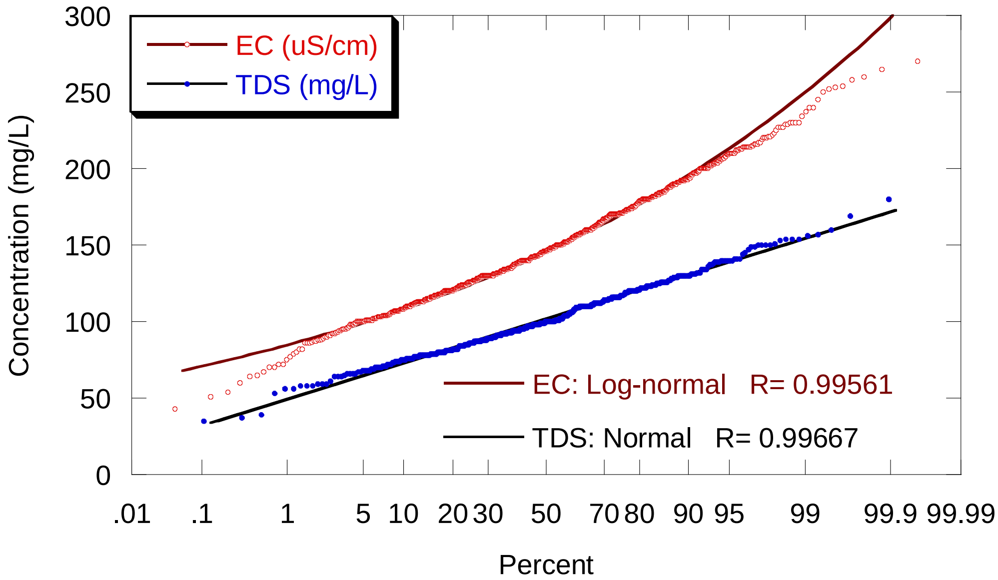

Figure 1: Example Probability Plot of Electrical Conductivity and Total Dissolved Solids in the Sacramento River at Freeport.

STATISTICAL REGRESSION ANALYSIS

Data analysis was performed using Minitab, a statistical software program. Single and multiple regression analyses were performed to determine whether the ambient data appear to be affected by time, season, and/or river flow (where available/applicable). For each constituent-location pair, regression analysis was used to generate an equation describing the statistical relationship between the predictor variables (time, season, and river flow), and the response variable (ambient concentration data). The regression generally used the ordinary least squares method which derives the equation by minimizing the sum of the squared residuals. From these equations, it was determined whether the ambient data show statistical trends with time, season, and/or river flow, the direction of the trend, and the statistical significance of each relationship.

INTERPRETING REGRESSION RESULTS

An example of the statistical output generated by Minitab is shown in **Table 1**. The presence of a trend is determined from the “p-value”. The p-value determines the appropriateness of rejecting the null hypothesis in a hypothesis test. In this case, the null hypothesis is that there is no trend. P-values range from 0 to 1. The smaller the p-value, the smaller the probability that rejecting the null hypothesis is a mistake, or the greater the probability that a trend does exist. For this analysis, a limiting p-value of 0.05 was chosen. If the p-value of the analysis was less than 0.05, the null hypothesis was rejected and a trend was assumed to exist. In the example provided in **Table 1**, the regression analysis was performed on concentration data versus the decimal year (time), Freeport flow, and season. In this case, all of the predictor variables hold p-values of 0.000, much less than 0.05, and therefore trends were assumed to exist with time, flow, and season.

Table 1: Example of Minitab Regression Analysis Results.

Regression Analysis: Concentration versus Decimal Year, Freeport Flow, Season

The regression equation is:

LN(concentration) = 29.4 – 0.0159 Decimal Year – 0.000004 Flow + 0.161 Season

Predictor	Coefficient	p-Value
Constant	29.37	0.000
Decimal Year	-0.015908	0.000
Freeport Flow	-0.00000364	0.000
Season	0.16054	0.000

R-Squared = 22.4%

R-Squared (adjusted) = 22.0%

The positive or negative value of the coefficient assigned to each predictor indicates the direction the trend takes. In the example shown in **Table 1**, a negative coefficient assigned to the Decimal Year predictor indicates that concentration data is likely to have a downward trend with time. There is an apparent downward trend with flow as well, and an upward trend with season as it progresses from summer to winter. The "R-squared (adjusted)" value was used to determine the goodness-of-fit of the regression line with the data. R-squared is the percentage of concentration data variation that is explained by its relationship with one or more predictor variables (time, season, and/or flow), adjusted for the number of predictors in the model. This adjustment is important because the R-squared for any model will always increase when a new predictor is added. A model with more predictors may appear to have a better fit simply because it has more predictors. However, some increases in R-squared may be due to chance alone.

Regression analyses performed on data in the Sacramento River (where flow data were available, allowing three predictors: time, season, and flow) resulted in adjusted R-squared values ranging from 3.6% to 80.7%, with a median of 24.9% and standard deviation of 17.5%. Regression analyses performed on data collected at locations not on the Sacramento River (where only two predictors were available: time and season) resulted in adjusted R-squared values ranging from 0.0% to 81.3%, with a median of 19.5% and a standard deviation of 16%. An adjusted R-squared value greater than the median plus standard deviation was considered to indicate an "excellent" fit with the data; a value greater than the median was considered to indicate a "good" fit; a value greater than the median minus the standard deviation was considered to indicate a "fair" fit; and a value below that was considered to indicate a "poor" fit with the data. Therefore, the adjusted R-squared value of 22.0% shown in **Table 1** (where three predictors were available) was considered to indicate a "fair" goodness-of-fit with the data.

DESCRIPTION OF RESULTS

The results of the trend analysis effort were compiled in tables and displayed in time series graphs, and are presented in the individual constituent evaluations for Category 1 constituents. Time series graphs were developed showing actual ambient concentrations beside the concentrations predicted by the regression equations. A best-fit line through the predicted concentrations visually shows the direction of the trend, where a trend is statistically shown to exist. An example of these time series plots is shown in **Figure 2**.

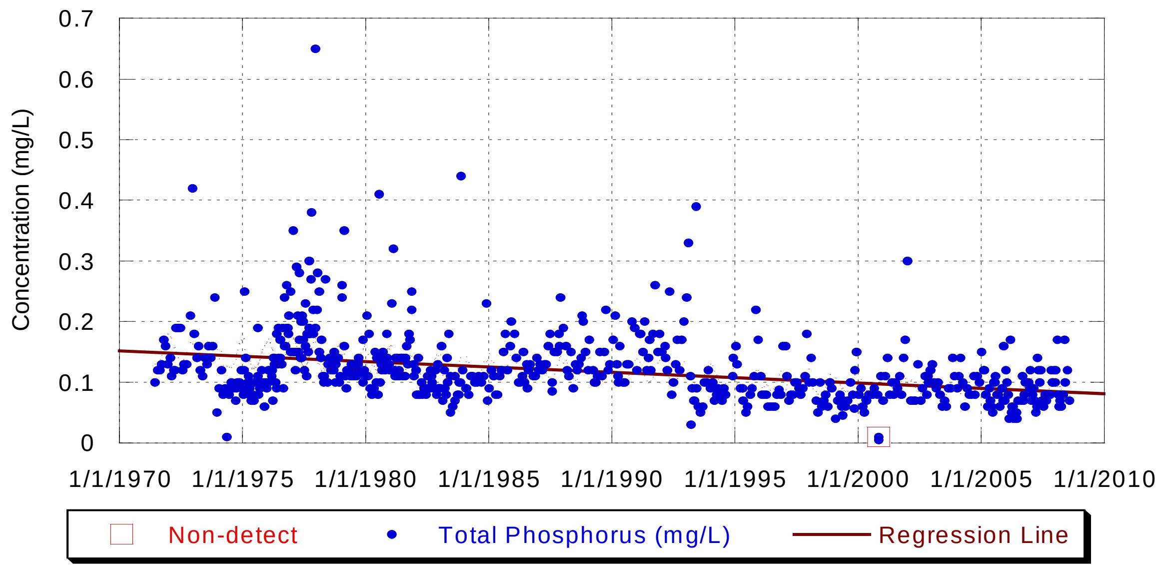

Figure 2: Example of Historic Data and Regression Analysis Trend Line for Total Phosphorus in the Sacramento River at Greene's Landing/Hood.

DATE RANGES OF AMBIENT DATA

Table 2 summarizes the date ranges of the water quality concentration data compiled for each trend analysis constituent at each location. The earliest water quality data were from 1958, the most recent from 2008. Daily river flow data from October 1948 to October 2008 were compiled from the USGS National Water Information System II Web Interface and used in the trend analysis where applicable.

CONCLUSION

Trend analyses were performed for Category 1 constituents at various near-field (Sacramento River at Freeport) and far-field locations as a means to determine whether their concentrations have significantly changed over time in the Sacramento River and Delta. Delta locations of interest for which trend analyses were performed include drinking water supply intakes. Where available, ambient water quality data collected as far back as 1958 and as recent as 2008 were used for trend analysis. A multiple regression analysis was performed to determine whether trends exist between pollutant concentrations and time, flow, and/or season. The analysis showed that for many constituents either downward or upward trends between concentration and time do exist for the Sacramento River and Delta locations evaluated. The results of the trend analyses are presented in the individual pollutant evaluations for Category 1 constituents included in the antidegradation analysis.

Table 2: Date Ranges of Data used in Regional Surface Waters Trend Analysis.

	Date Range of Trend Analysis Data by Location					
Constituent	**Freeport**	**Greene's Landing/ Hood**	**Emmaton**	**CCWD PP #1**	**CCWD Los Vaqueros Intake**	**Banks Delta PP**
Ammonia as N	1979-2008	1979-2008		1996-2008	1996-2008	1991-2008
Total Nitrogen	1973-2008	1974-2008		2002-2008	2002-2008	1998-2008
Nitrate as N(1)	1958-2008	1972-2008		1991-2008	1990-2008	1990-2008
Total Kjeldahl nitrogen	1973-2008	1972-2008		2002-2008	2002-2008	1997-2008
Total phosphorus	1970-2008	1971-2008		2002-2008	2002-2008	1997-2008
Electrical conductivity	1958-2008	1971-2008	1988-2000	1990-2008	1989-2008	1982-2008
Chloride	1958-2008	1971-2008		1990-2008	1989-2008	1982-2008
Total organic carbon	1973-2008	1983-2008		1996-2008	1996-2008	1986-2008
Total organic carbon	1973-2008	1983-2008		1996-2008	1996-2008	1986-2008

(1) Since nitrate comprises the major portion of a nitrate plus nitrite total concentration, ambient nitrate data were used in the trend analysis for the determination of trends in nitrate plus nitrite concentrations in the lower Sacramento River and Delta in the absence of sufficient nitrate plus nitrite data.